DOCUMENT RESUME

ED 273 022 EA 018 756

AUTHOR	Ross, Doris; Solomon, Lester
TITLE	Evaluating Teachers: With Lessons from Georgia's Performance-Based Certification Program.
INSTITUTION	Education Commission of the States, Denver, Colo.
SPONS AGENCY	Department of Education, Washington, DC.; Spencer Foundation, Chicago, Ill.
REPORT NO	ECS-TQ84-2
PUB DATE	Jul 85
GRANT	G00-830-3600
NOTE	26p.
PUB TYPE	Information Analyses (070)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; Observation; Performance; State Programs; Teacher Certification; *Teacher Evaluation; Teacher Supervision; Testing
IDENTIFIERS	Georgia

ABSTRACT

An overview of the reasons for and basic elements of teacher evaluation programs, and a deeper look at one state's program in particular, are presented in this booklet. The first chapter discusses why teacher evaluations are useful, what processes for evaluation can be followed, how state and local policymakers can enhance the usefulness of teacher evaluation, and why evaluation programs must be continuously updated. Chapter 2 briefly describes evaluation systems used in five states: an internship program in Kansas, career ladders in Arizona and Tennessee, Mississippi's Accountability/Instructional Model, and New Mexico's Staff Accountability Plan. Georgia's teacher evaluation program is described in detail in chapter 3. Among the aspects of Georgia's program that are covered are teacher testing in basic skills, general knowledge, professional education, and specific subject areas; the use of norm- and criterion-referenced tests; certification and testing; on-the-job assessment; the selection and training of evaluators; the scheduling of evaluations; and staff development efforts aimed at meeting demonstrated needs. Chapter 4 identifies factors to consider when evaluating for promotions or salary increases. Twelve references and a list of nine bookets available from the Education Commission of the States are provided. (PGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED273022

EVALUATING TEACHERS

With Lessons From Georgia's Performance-Based Certification Program

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EA 018 756

EVALUATING TEACHERS

With Lessons From Georgia's Performance-Based Certification Program

No. TQ84-2

By
Doris Ross
Education Commission of the States
and
Lester Solomon
Georgia Department of Education

Education Commission of the States
1860 Lincoln Street, Suite 300
Denver, Colorado 80295

July 1985

Work on this paper was supported by funds from the Spencer Foundation, the Education Commission of the States and the U.S. Department of Education, Grant no. G00-830-3600.

The authors would especially like to thank Ellen Fiannelly, Danette Padia and Anna Likens for their important contributions to the completion of this paper.

The views expressed in this paper are the authors' and do not necessarily reflect the views of the Education Commission of the States.

The Education Commission of the States is a nonprofit, nationwide interstate compact formed in 1965. The primary purpose of the commission is to assist governors, state legislators, state education officials and others to develop policies to improve the quality of education at all levels. Forty-eight states, the District of Columbia, American Samoa, Puerto Rico and the Virgin Islands are members. The ECS central offices are at 1860 Lincoln Street, Suite 300, Denver, Colorado 80295. The Washington office is in the Hall of the States, 444 North Capitol Street, N.W., Suite 248, Washington, D.C. 20001.

It is the policy of the Education Commission of the States to take affirmative action to prevent discrimination in its policies, programs and employment practices.

CONTENTS

INTRODUCTION: WHAT IS EVALUATION?

1. OVERVIEW OF TEACHER EVALUATION 5
Why Evaluate Teachers?
How Will Evaluation Information Be Used?
What Means Are Used To Evaluate?
The Policy-Maker's Role
The Evaluation Cycle: Never Ending

2. STATE APPROACHES TO EVALUATION 10
The Kansas Internship Program
Arizona Career-Ladder Pilot Project
Career Ladders in Tennessee
The Mississippi Accountability/Instructional Model (AIM)
New Mexico's Staff Accountability Plan
Georgia

3. LESSONS FROM GEORGIA'S PERFORMANCE-BASED CERTIFICATION PLAN 12
Teacher Testing
Norm- and Criterion-Referenced Tests
Georgia's Testing Program
Setting Minimum Scores
Implementing Testing Programs
On-the-Job Assessment
Tools for Observation
Who Should Evaluate?
Georgia's Assessment Process
Training
Scheduling
Performance Profile
Staff Development

4. EVALUATION FOR PROMOTIONS AND SALARY INCREASES 20

REFERENCES

INTRODUCTION: WHAT IS EVALUATION?

Formal evaluation is a term that makes most of us a little bit nervous. Yet, nearly every time we make a decision, we have first evaluated our choices. When we buy a new pair of shoes, we first "evaluate" several pair. Are they suitable for the use we intend for them? Do they fit or do they offer a hint of a "pinch" that may become intolerable? Are they comfortable? Do we like their style? Does the construction have any weak spots that may cause trouble later on? Can we afford them? The pair of shoes that meets most of our "criteria" is generally the pair that ends up on our feet. This is a simple example of informal evaluation.

We tend to select our friends and associates through an informal evaluation process as well, but, because we are dealing with human beings, not inanimate objects, it is a more serious decision, with implications that are important to our lives and to the lives of those with whom we associate.

Formal evaluation of individuals or programs that serve people is the toughest process of all, because the "ripples" created by evaluation-based decisions will spread from the individual being evaluated, to the program the person is involved in, to the institution providing the program, to the community.

When applied to people, it is a task that cannot be taken lightly. It is undeniably a subjective process that is essentially comparative. According to Ernest House in **Evaluating with Validity,** it "persuades rather than convinces, argues rather than demonstrates, is credible rather than certain, is variably accepted rather than compelling."

The nebulous character of evaluation in general, and formal evaluation in particular, argues for thoughtful criteria and procedures that will produce the best and most accurate judgment for decision making.

1. OVERVIEW OF TEACHER EVALUATION

Why Evaluate Teachers?

The push for education reform in the eighties has, not surprisingly, produced state and/or local mandates for evaluation of teachers. In general, the reasons behind these mandates are based on a desire for better educated students. Simple deduction tells us that students will be better educated if they are better taught. To improve teaching, we first need to find out how well it is being done. We need to identify teacher strengths and weaknesses and then decide what action can be taken to improve teachers and teaching skills.

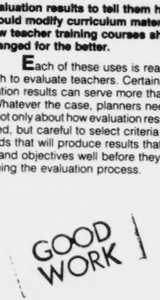
Again: Our goal is to better educate students. States and the public have chosen to evaluate teachers, identify appropriate skills to work on and then to take corrective or positive action.

How Will Evaluation Information Be Used?

Any policy maker or administrator contemplating a teacher evaluation program must know, first of all, how evaluation results will be used. This kind of backward-mapping will ultimately define and shape the process. Evaluation policy makers and planners may want:

1. **To provide teachers with feedback about their performance, with information they can use to improve their performance. In an extension of this, they may want to design and offer inservice training programs to teachers who fall short of the standards they have set.**
2. **To use evaluation results as one indicator of whether or not a teacher candidate is ready to be certified, or to decide whether a practicing teacher's certificate is to be renewed.**
3. **To use evaluation results to help decide whether teachers should receive salary raises or be promoted to a higher grade. On the opposite side, they may want to identify teachers to dismiss.**
4. **Evaluation results to tell them how they should modify curriculum materials, or how teacher training courses should be changed for the better.**

Each of these uses is reason enough to evaluate teachers. Certainly, evaluation results can serve more than one use. Whatever the case, planners need to be clear not only about how evaluation results will be used, but careful to select criteria and methods that will produce results that match goals and objectives well before they begin designing the evaluation process.



What Means Are Used To Evaluate?

Student testing. The general public has already evaluated our teachers; this is one underlying factor in the call for education reform. One way the public evaluates teachers is by looking at student test results — assessment results. If scores are low, teaching must be bad. But this route to teacher evaluation is a dangerous one, for, as a single measure, its results are questionable. Teachers do not work under the same conditions; their students are from varying socioeconomic, cultural and ethnic backgrounds. Results will be unfairly skewed by these and other factors. If student test scores are to be used, they should not be used alone, but as only one part of a larger evaluation scheme.

Testing teachers. Fairness is crucial to any evaluation process. If student test results are not a fair way to evaluate, what about testing the teachers? Certainly, this is fairer. However, it is also a difficult challenge. The test design is of paramount importance. No matter how hard planners and designers try to make such tests produce accurate measurements of teacher knowledge, they cannot avoid some subjectivity. Tests are designed by human beings with attitudes and biases that may creep in, in spite of efforts to avoid them. This does not suggest, however, that tests cannot be used as an evaluation tool. Indeed, they are often the best way to gather essential **quantitative** data. But such tests alone may not provide enough information for decision making. If tests are to be used, planners need to be alert to the need for very careful design that does not ultimately discriminate against some teachers who take the test. They need to be sure, too, that the tests measure only the knowledge and skills appropriate for successful performance of the teaching job. Further, minimum scores need to be set at realistic levels. Finally — and perhaps most important — planners need to place test results into a larger evaluation context, and to assign them a value that is commensurate with their best and most careful perception of their validity.

Observing teachers. The subjective, comparative nature of evaluation, understood and creatively applied, can be a distinct asset for the evaluation process. While tests can provide reasonably reliable quantitative data (How much of the subject does the teacher know?), when planners consider making an important "quality" decision, they typically find that quantitative data is not enough. They often will turn to qualified "experts" for their observations. For teacher evaluation, the experts can be

- **Other teachers**
- **Principals or departmental supervisors**
- **Professional evaluators**
- **Students and parents**

Each of these experts, regardless of his or her expertise, will need to understand fully the purposes of the evaluation, and be carefully trained to evaluate the subject according to predetermined criteria and procedures. Still, each will bring a different — and valuable — set of perceptions to the process. Their subjectivities, when combined, can broaden and smooth out the data gathered for decision making. Consulting more than one expert, then, is likely to result in evaluations that are as fair and as accurate as they can be. This will not only humanize the process, it will move results closer toward what is essentially unobtainable as an absolute: objectivity.

Evaluation by observation will either stand alone as an evaluation tool or can be paired with testing as a basis for decision making. However, if evaluation by observation is conducted by unskilled and untrained evaluators who are working with criteria and standards that do not fit the ways the evaluation information is to be used, the results will be useless and the effort will be divisive rather than helpful. In designing a process for evaluation by observation, and assuming that planners know why they want to evaluate and how they will use the information, they need to ask — and answer — questions like these:

1. **Who will evaluate?** As noted above, several choices or combinations of choices can be considered. An obvious evaluator is the teacher's supervisor — the building principal. There is peer evaluation — appraisal by the teacher's colleagues in his or her school. If practicing principals and teachers are to be evaluators, can planners arrange released time for them and supply substitutes? If planners feel these kinds of evaluators are "too close to home," with possible unhappy results of strained teacher relationships (and increased teacher isolation), in a school, they may want to consider external evaluation as well.

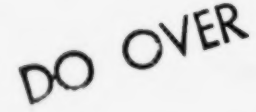
   External evaluators may be principals and/or teachers from other schools, district-level or state-level education agency personnel or even outside consultants not directly connected to the public education system. Certainly, external evaluators may be more objective. But there is also the danger that they may be too far removed from the situation to produce an evaluation that will stand alone.

   As implied above, the fairest and most reliable approach may combine the observations of several kinds of evaluators. **Those chosen to evaluate must be carefully trained; they must understand the reasons for evaluation and how the process will generate usable information.**

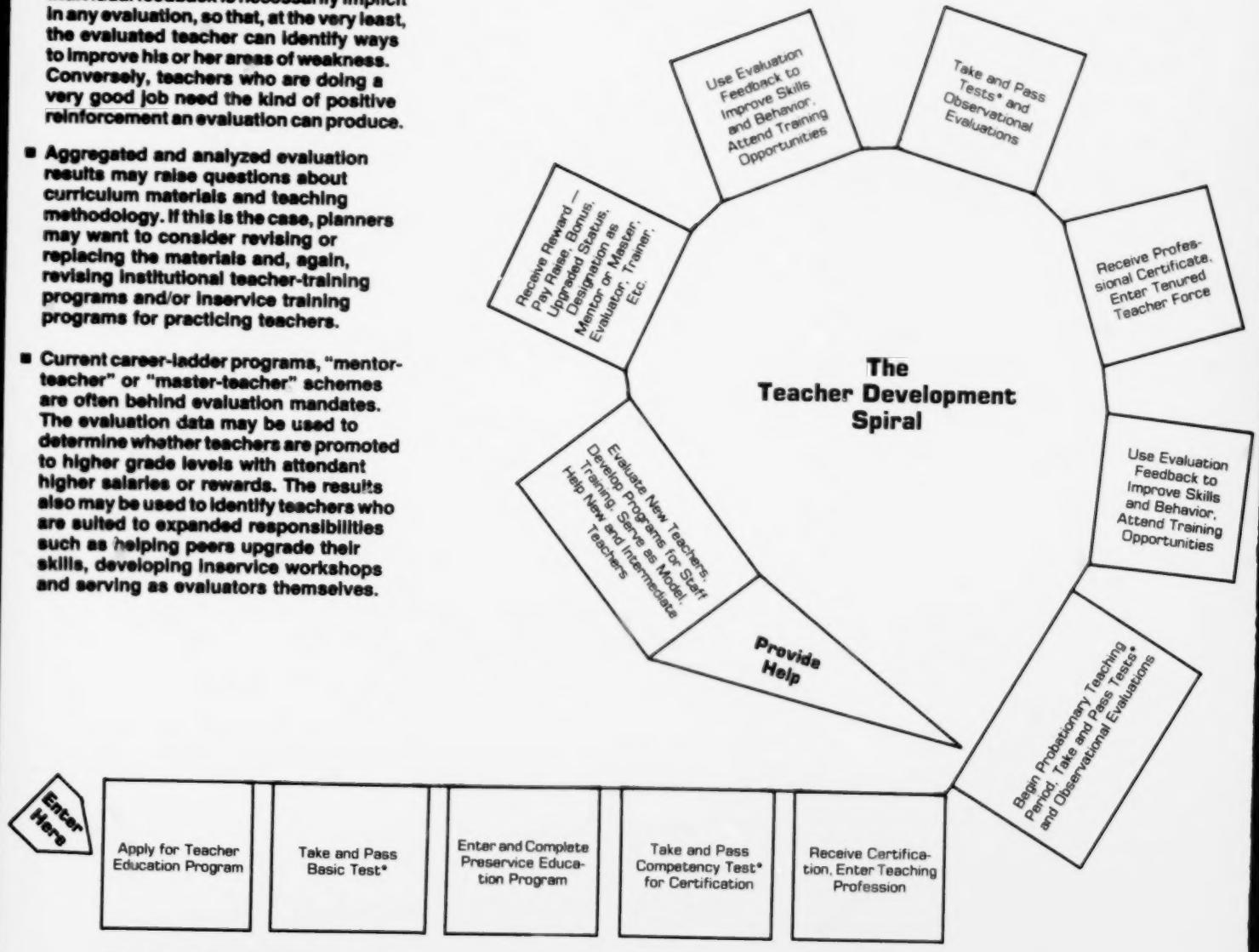
2. **How often should evaluation be done?** Evaluating every teacher every year seems like a good idea, but it may be too expensive or not feasible for other reasons. Should teachers go through the cycle every two or three years? Should less-experienced teachers be evaluated more often than more-experienced teachers? Should teachers at the top of the salary or grade heap be exempted from evaluation? If planners choose to evaluate every three years, for instance, will they cover all of the teachers in a year's time or will they spread the evaluation over a three-year period, ultimately covering every teacher?

Using evaluation results. Each time a selected group of teachers has been evaluated, the information must be analyzed and fed to those who will act on it: individuals, a school, a district, teacher-training institutions or state administrators and policy makers.

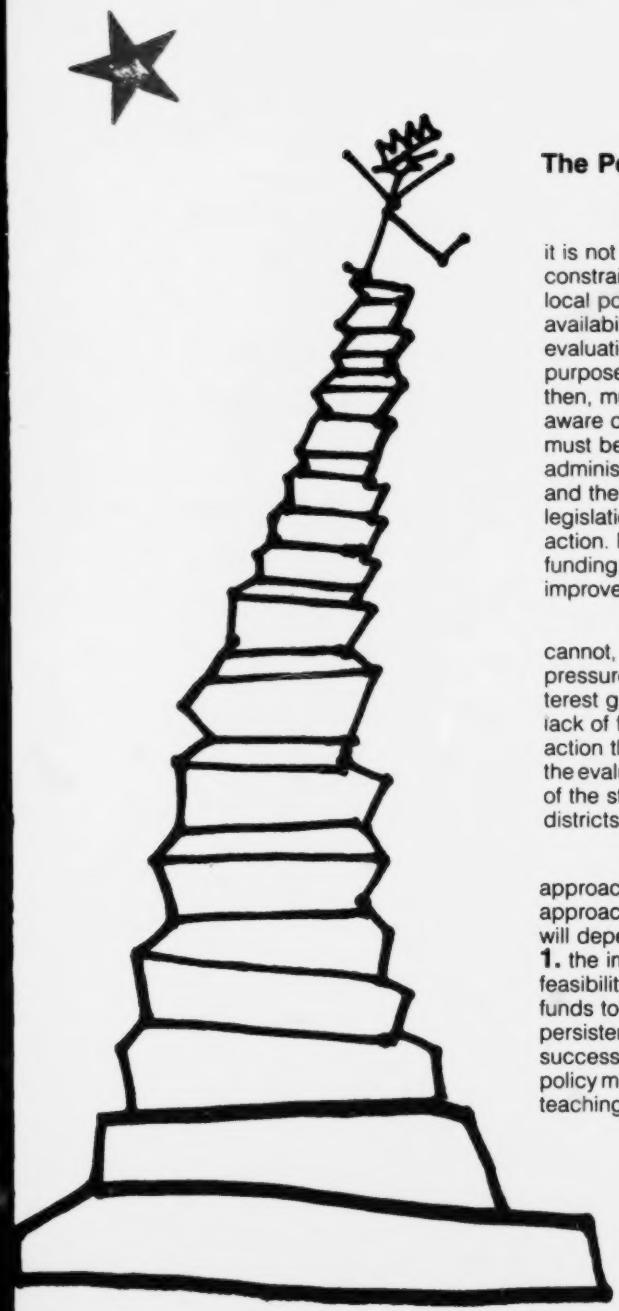
- **If evaluation results are to help planners decide whether or not teacher candidates are qualified for certification, or whether practicing teachers are performing well enough to continue to be certified, they need to be analyzed individually and in the aggregate to see if they meet certification criteria. If too many individual teachers fall short of certification criteria or if evaluation results do not "fit" the criteria, planners and implementers may want to consider 1. re-examining the criteria themselves, 2. re-examining the evaluation process or 3. revising teacher-training programs in higher education institutions. Practicing teachers who do not measure up may need inservice training programs to beef up their teaching skills.**



- **Certainly, the teachers who have been evaluated need to know how they did. Individual feedback is necessarily implicit in any evaluation, so that, at the very least, the evaluated teacher can identify ways to improve his or her areas of weakness. Conversely, teachers who are doing a very good job need the kind of positive reinforcement an evaluation can produce.**
- **Aggregated and analyzed evaluation results may raise questions about curriculum materials and teaching methodology. If this is the case, planners may want to consider revising or replacing the materials and, again, revising institutional teacher-training programs and/or inservice training programs for practicing teachers.**
- **Current career-ladder programs, "mentor-teacher" or "master-teacher" schemes are often behind evaluation mandates. The evaluation data may be used to determine whether teachers are promoted to higher grade levels with attendant higher salaries or rewards. The results also may be used to identify teachers who are suited to expanded responsibilities such as helping peers upgrade their skills, developing inservice workshops and serving as evaluators themselves.**


The Teacher Development Spiral
Enter Here
Apply for Teacher Education Program
Take and Pass Basic Test*
Enter and Complete Preservice Education Program
Take and Pass Competency Test* for Certification
Receive Certification, Enter Teaching Profession
Begin Probationary Teaching Period, Take and Pass Tests* and Observational Evaluations
Use Evaluation Feedback to Improve Skills and Behavior, Attend Training Opportunities
Receive Professional Certificate, Enter Tenured Teacher Force
Take and Pass Tests* and Observational Evaluations
Use Evaluation Feedback to Improve Skills and Behavior, Attend Training Opportunities
Receive Reward — Pay Raise, Bonus, Upgraded Status, Designation as Mentor or Master Evaluator, Trainer, Etc.
Evaluate New Teachers, Develop Programs for Staff Training, Serve as Model, Help New and Intermediate Teachers
Provide Help

*This booklet discusses the kinds of tests that can be used in teacher evaluation.



The Policy-Maker's Role

Unless evaluation results in action, it is not worth doing. If action is prohibited, constrained or sidetracked by state and/or local policies, legislation, regulation and/or availability of funds and funding practices, evaluation information serves no useful purpose. State and local policy makers, then, must be "enablers." They must be aware of aggregate evaluation results, they must be responsive to recommendations of administrators and evaluation personnel, and they must promote policies and legislation that will set the stage for effective action. Further, they should work to identify funding sources that will support action for improved teaching.

Most often, state policy makers cannot, because of myriad political pressures — from colleagues, special-interest groups, the general public — and/or lack of funds, set the stage for **sweeping** action that will address all of the problems the evaluation process has identified. Many of the steps must be taken by individual districts and schools.

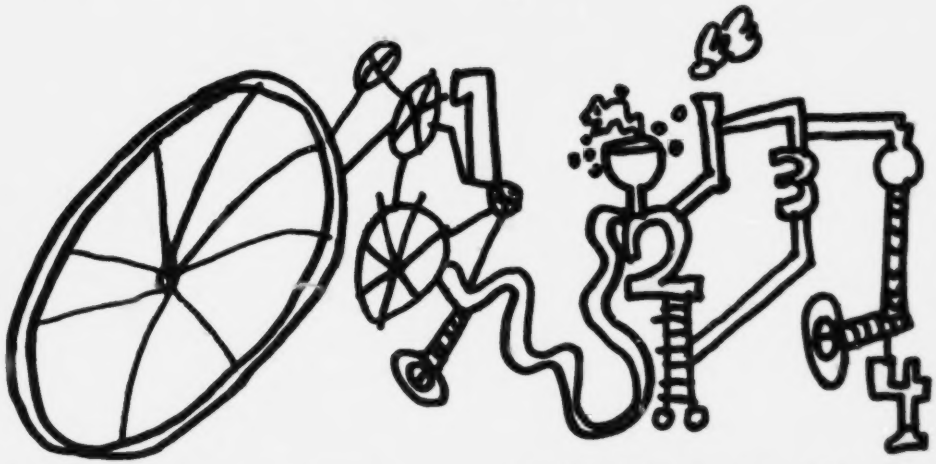
Incremental steps are a traditional approach to change. In an incremental approach, which stage-setting steps are taken will depend on one or more of these factors: **1.** the importance of the step, **2.** the political feasibility of the step and **3.** the availability of funds to implement the step. With some persistence over time, and with measurable successes in the steps along the way, the state policy maker can be instrumental in improving teaching.

The Evaluation Cycle: Never Ending

It would be gratifying indeed if the teacher-evaluation cycle described in these pages could be wrapped up, once and for all, in a single run-through. Certainly, once around the track is better than no trip at all. However, this is an unrealistic expectation.

Any new program is not perfect, and when it is put into place, "wrinkles" will surface. Standards and criteria may need to be altered; facets of the process may prove to be impractical or, indeed, unworkable; institutional and inservice training programs may need to be adjusted to produce better results; and overall expectations for the process may not be met. Further, an ever-changing societal, cultural, political and economic context will produce new demands and new expectations on teachers and teaching. If goals, objectives, standards, criteria and the evaluation process itself remain static, the result can only be another "crisis," another call for reform at a later date.

2. STATE APPROACHES TO EVALUATION



Many states are either considering new teacher evaluation schemes or are in the process of putting them into place. No two states have chosen the same system. The following paragraphs briefly describe **some** state activity. These examples were not selected because they are the "best," but merely to provide an overview of the kinds of evaluation schemes being implemented. Time and circumstances will eventually reveal what is "best" — and even then, what is "best" for one state or district may not be "best" for another.

The Kansas Internship Program

Late in 1981, the Kansas State Board of Education set wheels in motion to develop an internship program for new teachers. After considerable work, the legislature in 1985 approved plans for the program and appropriated first-year funds.

The plan calls for locally selected "assistance committees" of professional educators (including senior teachers and a certified administrator) to assist, support and assess (evaluate) all first-year teachers. The evaluation component of the plan calls for the use of a single, standardized statewide assessment instrument (test) to measure first-year teacher performance. This measure will be supported by observational evaluation and conferencing (feedback). Committee members will be carefully trained in all of the evaluation techniques.

The program will be installed in four phases:

- **Phase I beginning July 1, 1985. The state department of education develops an assessment instrument and designs training programs for assistance and assessment.**
- **Phase II beginning July 1, 1986. Following regional meetings and technical assistance to local education agencies, districts to conduct pilot projects will be selected, the assessment instrument will be validated, a statewide assessment will identify needs of first-year teachers, and colleges and universities will submit applications to conduct seminars to meet those needs.**
- **Phase III beginning July 1, 1987. With full state funding support, districts will begin field-testing the program, and a data base will be developed.**
- **Phase IV beginning July 1, 1989. The full program will be in place and the state will begin to issue one-year intern certificates.**

For more information, contact Diane Castle, Kansas Department of Education, 913-296-3108.

Arizona Career-Ladder Pilot Project

On June 1, 1985, 16 school districts in Arizona submitted plans for local career-ladder plans for teachers. These plans were developed with state planning grants: one additional district developed plans on its own. Under state legislation passed in 1984, district plans "must be based on a completely restructured salary schedule in which a salary range is set for each level on the career ladder and the salary for a teacher within the range is based on **objective performance evaluation** or other objective factors." For evaluation, final plans must describe the criteria for advancement to each career-ladder level and evaluation procedures that include **more than one measure of performance,** one of which must be a measurement of teacher performance in relation to student academic progress.

State payments for career-ladder plans will be allocated through an increase in the district's per-pupil block grants. Fiscal year 1985–86, designated as a training and evaluation year, will see a .5% increase for selected districts. Subsidies will top out in 1989–90 at 5%.

For more information, contact Judy Richardson, Arizona Department of Education, 602-255-3171.

Career Ladders in Tennessee

Tennessee's career-ladder program for teachers, funded in 1984 and implemented that summer, is a state-level effort. Steps in the ladder (probationary, apprentice and career levels I, II and III), timelines, evaluation criteria, procedures, salary increments, and so forth have been set by state law and policy. Teacher participation is optional. The evaluation component begins for practicing teachers who are qualified with a specific number of years of experience and who wish to "fast-track" onto the ladder, with the **National Teachers' Examination.** Teachers qualify for each step of the ladder by "passing" a **professional skills test,** by "passing" an observational evaluation by an out-of-district team, by peer, student and principal evaluations, all of which are considered in recommending advancement to the state board of education. At the probationary level, evaluation is a local affair; as teachers move up, the state reviews local evaluations at three-year to five-year intervals.

For more information, contact Janet R. Handler, 615-974-2431 or Deborah L. Carlson, 615-974-4441, at the University of Tennessee.

The Mississippi Accountability/ Instructional Model (AIM)

AIM is part of the legislature's Education Reform Act of 1982, and its evaluation component weaves its way through six interactive areas in each school: leadership, organization, instruction, staff development, climate and pupil achievement. The goal is performance-based school accreditation.

Across-the-board standards developed by the state include some broad standards for teachers that call for some kind of evaluation process. In preparation for instituting a "personnel appraisal and compensation system," recommendations for pilot programs are being developed and will be reported to the legislature on January 5, 1986.

For more information, contact Jim Hancock, Mississippi Department of Education, 601-359-3513.

I SEE ME !

New Mexico's Staff Accountability Plan

In July 1986, school districts in New Mexico will be using a state-developed "staff accountability framework" to evaluate the performance of principals and teachers. The process will measure six essential competencies for both principals and teachers. Local evaluation plans submitted in July 1985 and developed in accordance with state guidelines must focus on the growth and development of the staff member. Each local plan must define competencies, include descriptors, provide for training participants in observation techniques, conference skills and growth planning and present a schedule for training and implementation.

For more information, contact Paula Gottleib, New Mexico Department of Education, 505-827-6573.

Georgia

Georgia's teacher evaluation program is described in some detail in Part 3 of this booklet. The program was selected for illustration because

- **It has been in place for five years and thus has a credible experience base.**
- **It contains many of the components discussed in Part 1 of this booklet.**

In the following pages, Lester Solomon, with career-ladder mandates in mind, describes both the Georgia program and the rationale behind its development.

3. LESSONS FROM GEORGIA'S PERFORMANCE- BASED CERTIFICATION PLAN

by Lester Solomon

Although Georgia's performance-based certification program is frequently thought of as a stepping stone to a career-ladder program, it was initiated primarily to guarantee the competency of beginning teachers. Assessment of beginning teachers was begun in 1980, after nearly 10 years of planning, development, field-testing and training. In Georgia, we have learned that obstacles, however formidable, can be overcome and we have learned something about what works and what does not. These lessons can be applied to evaluating experienced teachers for career ladders as well as evaluating beginning teachers.

Teacher Testing

Types of tests. In determining how much a teacher "knows," paper-and-pencil testing is invariably the first, and too often the only, method considered. Testing should reveal what teachers know, when they should know it and how the knowledge being measured relates to teaching competence. Possible areas for testing include basic skills, general knowledge, professional knowledge and subject matter. As will be revealed below, considerable experience with paper-and-pencil testing over the past decade has convinced us in Georgia that the only appropriate test to be used for a part of an incentive program such as a career ladder is the subject-matter test.

Basic skills. Most people agree that reading comprehension, mathematical computation and writing skills should be tested. What is controversial is when this type of testing is best done. The strongest case can be made for basic-skills testing as an entrance requirement to teacher education. Passing this type of test provides some assurance that students are ready to enter a teacher-training program, and that they can profit from the instruction.

Basic skills testing at the point of college graduation, initial certification or for career-ladder eligibility may be "too little, too late." At these points, tests for subject-matter knowledge will almost invariably reveal whether or not those being tested have the necessary job-related basic skills. Those who slip through early basic-skills screening are not likely to get through a second, but more indirect, screening.

EXCELLENT WORK

General knowledge. The purpose of testing is to identify those who are qualified to teach. Test content must be job-related. Teachers' general knowledge relates very little to actual job performance. Further, general-knowledge tests tend to have the greatest adverse impact on minorities. Using these tests as a factor in placing teachers on career ladders may have undesirable and possibly illegal effects; teachers populating the upper rungs of the ladder may not include appropriate numbers of minorities.

Professional education. While testing how well teachers have learned the art of teaching seems reasonable, the kinds of questions that typically appear on professional education tests tend to emphasize the philosophical foundations of teaching, rather than skills that teachers actually utilize in the classroom. Understanding the foundations of teaching is an important part of teachers' education, but it cannot be **directly** related to the requirements for advancing in a career ladder program.

In Georgia, evidence convinced us that direct observations of teachers in the classroom will provide the best index to how effective teachers are. To make our classroom evaluation of teachers' professional knowledge as thorough as possible, our on-the-job evaluation includes assessing teaching plans and materials.

Subject area. Teachers should know the content of the field in which they are certified and employed to teach. Already-developed teaching-field tests, highly job-related, are available, which have been validated for both essential subject knowledge and the amount of time spent teaching various aspects of the content. However, because this type of test does not determine whether an individual is a good teacher, scores well above predetermined "mastery levels" should not be given undue weight. There is no evidence that knowledge of the subject beyond mastery levels relates to teaching effectiveness.

Knowledge of the subject matter, then, measured by validated subject-area tests, is a minimum requirement for effective teaching but is only one ingredient.

Test results that identify subject-area strengths and weaknesses can provide important information for employment selection, employment placement, continued self- and staff-development and graduate education. In employment placement, for example, where two applicants apply for a science position primarily involving teaching physics, criterion-referenced test results could report that one made a total score of 85 in science with a subarea score in physics of 75, while the other applicant made a total score of 80 with a 90 in physics. All other factors being equal, who should be placed in the physics teaching position?

Norm- and Criterion-Referenced Tests

Norm-referenced testing measures an individual's knowledge against a group "norm" or average. Criterion-referenced testing is based on carefully developed standards that do not relate to a "norm," but to particular areas of knowledge. As state planners design their evaluation programs, they will want to consider each type of test carefully, and to select the type that best meets their goals, objectives and criteria.

Most norm-referenced tests fail to provide clear expectations and feedback of meaningful results. If the objectives that teachers are going to be tested on cannot be identified, then that type of test should not be used for either teacher certification or career-ladder promotions. In Georgia, we have agreed that criterion-referenced testing has many advantages over norm-referenced testing. One of the basic rules in evaluation is to make expectations clear to people; the standards in criterion-referenced tests are firm and will not slip up and down with a group "norm." Teachers need to know what is going to count towards certification or advancement to upper-career levels.

OVER 27,000 TEACHERS

Evaluation should not be a process where a premium is placed on where teachers start but rather on where they finish. If the testing program's goal is to improve teachers' skills — not simply to weed out a few gross incompetents — it is vitally important that people who do not meet particular mastery levels get meaningful feedback and opportunities to improve their performance.

Criterion-referenced tests, properly focused, can provide detailed feedback identifying each applicant's areas of strength and weakness. Where these tests identify specific deficiencies, applicants can refer to the objectives associated with those areas as they prepare to retake the test. Study guides can provide references for review of subarea content and/or objectives.

Criterion-referenced tests, by identifying deficiencies and relating them to expectations, can result in success for the majority of those retaking them. Georgia, which has tested over 27,000 teachers from 1978 to 1984, reports a retake pass rate of 67% on criterion-referenced tests compared to about 33% or less in some states using the National Teachers' Examination, a norm-referenced test. Criterion-referenced testing programs, like those developed in Georgia, Oklahoma, Alabama and West Virginia and being developed in Texas, provide clear expectations to candidates through "published and public" objectives and clear feedback of results.

Georgia's Testing Program

The Georgia Teacher Certification Tests were developed and validated to be job-related and to reflect the minimum content knowledge necessary to teach in Georgia classrooms. The objectives and content of the examinations, as well as the minimum cutoff scores, were determined by committees of well-qualified Georgia educators from each certification field. Test items that measure the objectives were reviewed by the committees for item/objective content match, content accuracy, lack of bias and minimum competency. A job analysis, conducted throughout the state, determined how essential each objective was and how much teaching time was devoted to it. The items on each test exactly represent the specifications for the objectives and subareas as determined by the job analysis.

Setting Minimum Scores

Evidence does not support the proposition that the more teachers know, the more effective they are going to be. In Georgia, performance standards for teacher tests center on the minimum levels of performance necessary to demonstrate mastery of teaching-field content. "Standard error of measurement" (the range, above and below a specific score, in which measurement error is possible) associated with any test may be the best baseline for addressing concerns about minimums and maximums. To set cutoff scores that strengthen confidence in classification decision-making, we lowered validated minimum levels by 2½ standard errors of measurement. For tests with over 100 items, 2½ standard errors would constitute about 10%.

When testing teachers for upper levels of career ladders, this "fudge factor" — 10% on a 100-item test — could be eliminated to set a higher cutoff score, or added to set the cutoff score at the top of the range of the standard error of measurement. In this way, mastery levels can be raised (but not unrealistically) to provide increased confidence in the level of mastery achieved by the teachers being tested.

Implementing Testing Programs

In carrying out a testing program, the following components need to be considered.

- ***Registration.*** **Deadlines should provide teachers with enough lead time to prepare for test administrations and should allow enough time between test dates for teachers to receive their scores and prepare to retake the test if necessary.**
- ***Frequency of administration.*** **Tests should be given at least three times a year to provide applicants with repeated opportunities to take them. Many retake opportunities will assure due process. Records of retake success must be maintained.**
- ***Multiple sites.*** **Test-secure, standardized settings with trained proctors and reasonable access should be provided.**
- ***Cost of administration.*** **Costs are affected by frequency and number of sites (e.g., one administration per year at 1 site at a cost of $20 per applicant or six administrations a year at 12 sites at $60 are probably**

not as acceptable as three or four administrations per year at 4 or 5 sites at a cost of $35–$40 per applicant). Funds for item replacement and updating must be built into the cost of the examination.

- ***Test-score reporting.*** **Timely feedback must supply not only total scores but identify areas of strength and weakness.**

On-the-Job Assessment

Formal training does not necessarily produce a capable tradesman or professional. To promote quality and at least ensure minimum competence, most trades and professions require some form of licensure based on a test of competence. This is true of physicians, lawyers, plumbers, cosmetologists and many other trades or professions. In the past, it has not been true for teachers. Because colleges, education associations and local education agencies have been reluctant to deal with measures of competence, the responsibility ultimately falls to the licensing agency. As in most other states, the licensing agency in Georgia is the state board of education through the department of education.

The Georgia Performance-Based Certification Program, on the assumption that undergraduate teacher education cannot fully prepare a teacher to demonstrate professional competence in all the varied situations in the employment setting, extends preparation for teaching into the initial years of employment. An on-the-job assessment, coordinated by a network of 17 regional assessment centers, is required for all teachers who hold nonrenewable certificates.

This assessment focuses on teachers' abilities to organize, plan and conduct instruction as well as on classroom management and interpersonal relationships with students. The teachers are given three years (or six assessment opportunities) to demonstrate proficiency on all eight competencies necessary to earn a renewable certificate.

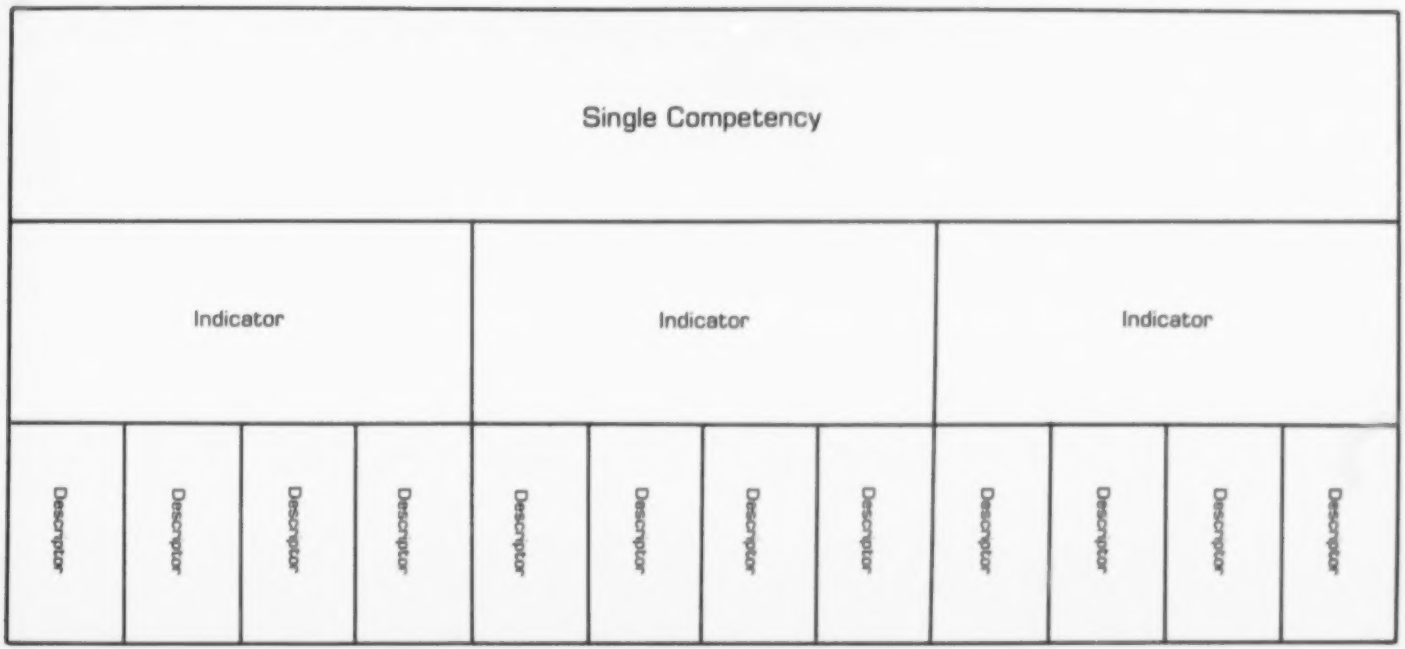
Thus, beginning teachers are assessed on the job on essential, generic teaching competencies. The local school system, in cooperation with others, provides these teachers with a continuous program of staff development, based on the needs identified in the on-the-job assessment.

Tools for Observation

Though not existent even 10 years ago, many well-tested descriptions of teacher competencies are available today. States and school districts can best spend time and energy determining how these competencies can be measured and how teachers can be trained to use them. Competencies must be essential and generic to teaching for a comprehensive and manageable assessment program. The job-relatedness of these competencies should be established across grade levels and subject areas. This can best be done by means of content-validity studies and review of research that relates the competencies to student outcomes. Major efforts should be focused on measuring the competencies, training evaluators and providing staff development opportunities.

Multiple tiers, with increasingly specific measurement items, are vital in on-the-job assessments. Focusing only on a single tier of behavior (e.g., the teacher communicates with learners . . .) results in assessment criteria that are either too broad and not directly measurable or too narrow to lend themselves readily to any type of organization. Even worse is an uneven approach that combines criteria that are too broad and too narrow. A two-tier approach is better but still lacks the precision necessary for ensuring reliable assessment and generating enough data for meaningful feedback. A three-tier approach assessing **1.** competencies as demonstrated using **2.** indicators and **3.** descriptors is best for an effective on-the-job assessment program. This multi-tier approach is also important for training observers and increasing their reliability.

Multiple Assessment Tiers for Teacher Evaluation



Detailed assessment tiers also help establish criterion-referenced standards. Many evaluation tools used in research do not lend themselves to either very good standard-setting or very good staff development. Evaluators wind up selecting norm-referenced approaches and requiring teachers to meet some kind of group average. On the other hand, if competency is measured by multiple indicators that have multiple descriptors, on-the-job assessments can be criterion-referenced for training staff and determining mastery.

Competencies. Some generic competencies can serve as the basis for on-the-job assessment of teachers. Teachers should:

1. **Plan instruction to achieve selected objectives**
2. **Obtain information about the needs and progress of learners**
3. **Demonstrate acceptable written and oral expression and knowledge of the subject**
4. **Organize time, space, materials and equipment for instruction**
5. **Communicate with learners**
6. **Demonstrate appropriate instructional methods**
7. **Maintain a positive learning climate**
8. **Maintain appropriate classroom behavior**

Indicators. After competencies have been determined, at least three to four indicators, **defining** the major behaviors necessary for competency, should be established to ensure reliable scores. These indicators should also be the basic units for reporting results.

Descriptors. Descriptors **identify** the specific observable behaviors related to each indicator, and are necessary to distinguish the quality of teacher performance.

Examples of specific teacher competencies with related indicators and descriptors follow.

Competency: Planning.
Plans instruction to achieve selected objectives.

Indicator (one of several). Specifies or selects learner objectives.

Descriptors:
1. Objectives are stated as performance outcomes.
2. Objectives describe an adequate scope/depth of content.
3. All learners have an adequate number of objectives that are appropriate to their instructional levels as delineated in the group description.
4. Different objectives are planned for learners with different instructional levels.

Competency: Observation.
Communicates with learners.

Indicator (one of several). Gives explanations related to lesson content.

Descriptors:
1. Explanations of lesson content are clear and easy to follow with appropriate vocabulary for learners.
2. Communication is precise with few false starts, qualifiers or interrupters.
3. Demonstrations and/or examples are used to illustrate content.
4. Major points or potential areas of difficulty are emphasized, using techniques such as repetition or verbal or nonverbal cues.

Detailed information about the development of teacher competencies is available from the Teacher Assessment Unit, Georgia Department of Education (404-656-2556).

Who Should Evaluate?

Considerable debate centers on who will assess teachers in a performance-based certification or career-ladder program. The strongest case can be made for an evaluation system of **multiple observations by multiple internal and external data collectors, including peer teachers.** Concern over the psychometric dimensions of reliability and validity often has overshadowed **credibility.** Peers and external people reinforce the credibility of an assessment system. Data will be more reliable when the entire team knows there is somebody external.

While administrators should not be the sole data collectors, they should not be excluded if there is to be any broad-based school improvement effort as a result of assessment. The key to involving administrators is extensive training and periodic reliability checks with external data collectors. Properly trained, administrators have better and more specific understandings of what to look for in the classroom, and many become enthusiastic supporters of on-the-job evaluations.

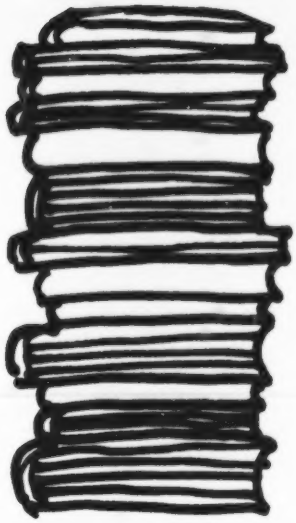
Peer teachers as data collectors help to balance an assessment system. Teachers being assessed may attach considerable credibility to evaluation by peers, particularly if they are certified in the same fields.

Often, peer evaluators themselves benefit from the process; they learn how to improve their own teaching. Even the district's best teachers, usually selected as evaluators, report learning much from paying detailed attention to discrete teacher competencies. Many return to their own classrooms with new ideas and renewed enthusiasm.

While states and school districts recognize the value of peer evaluators, they seem to be intimidated by the difficulty of including teachers in the process.

Georgia's Assessment Process

Georgia's **Teacher Performance Assessment Instruments (TPAI)** measure a teacher's performance on carefully selected and validated generic competencies. Different instruments measure: **1. Teaching Plans and Materials, 2. Classroom Procedures, 3. Interpersonal Skills, 4. Professional Standards** and **5. Student Perceptions.** Only the first three are applied for certification. The Professional Standards and Student Perceptions instruments provide bases for professional or staff development efforts. The former focuses on teacher relationships with colleagues and outside the classroom, acceptance of professional responsibilities and efforts to improve professional skills. The latter gauges how students react to some of the **same** indicators and descriptors assessed in the direct observation component.



Training

Each administrator, peer teacher or external data collector on an assessment team must be trained to use the Georgia **TPAI** and must meet proficiency requirements for the use of these instruments. The training includes 50 hours of instruction and practice with the **TPAI,** proficiency evaluations and field practice. Trainees rate videotaped teachers before they observe actual classrooms, and are required themselves to prepare portfolios and be assessed. Annual update training and interrater agreement checks help maintain the accuracy and skill of evaluators. State funds are provided to train and maintain all administrators and two to three teachers in every school in the state.

There is no need to wait until all details of the evaluation system are worked out to begin training people in the art of teacher observation. One of the best ways to start a ground swell of support for an incentive program is to begin, very early, training people to be teacher observers. By the time the evaluation system is put into place, people have been trained, have been going into classrooms and doing interrater checks on each other for so long that the process is second nature. However, training is not really complete until trainees get into real situations and establish reliability with trainers. As the process operates, continuing reliability checks on data collectors will maintain the system's quality.

Scheduling

Teachers up for evaluation are notified in advance, and specific classes and times are identified. The 7- to 10-day assessment period is based on the teacher's lesson plans, or portfolio. An assessment-team conference with the teacher is followed by three individual observations (a full class period each on three separate days) during the first 5 days, with the remaining days used for rescheduling if necessary.

Performance Profile

Georgia's regional assessment centers process all assessment data and develop confidential performance profiles for each teacher. Center personnel interpret these profiles in conferences with the teachers.

The profiles help local school systems (or designated agencies) plan staff development opportunities for assessed teachers. Performance data, also released to the college attended by the assessed teacher, provide information for program evaluation and improvement.

Staff Development

To change or improve teacher behaviors is a primary evaluation goal. The first step in bringing about change is letting teachers know what is expected of them. Traditionally, teachers begin their first jobs without knowing the competencies and indicators that should constitute their teaching repertoires. Without such information, they may become frustrated or develop poor teaching habits that are difficult to correct later. When expectations are clear, teachers have a foundation on which to develop competence, which is usually accompanied by improved self-assurance and ego-strength. A thorough orientation to the assessment process at the beginning of the school year is essential to communicate clear expectations.

Preservice training, induction and inservice training are truly points on a continuum; the development of competencies and assessment procedures should permeate the entire continuum. However, certain quantitative and qualitative distinctions between the expectations for student teachers during preservice training and teachers in the induction phase are necessary.

In Georgia, all assessable teachers are given the **TPAI** and may attend a follow-up orientation. This follow-up orientation raises teachers' understanding of what is expected of them in the assessment process from 75% to over 90%, and sets them on the route to positive teaching-behavior changes.

The on-the-job assessment, next, also stimulates change as teachers assemble their teaching plans and materials for their portfolios and demonstrate their classroom procedures and interpersonal skills during the course of three observations. In Georgia, evidence is clear that scheduled observations produce better performance in the classroom than random samples of teaching behavior. A teacher who "can do it" in a given instance has taken a first step toward consistently "doing it."

A more direct step toward behavioral change is providing specific feedback after an assessment. Assessed teachers receive interpretations of their performance profiles, which are focused on, but not limited to, established mastery levels. Over 85% of the beginning teachers assessed in Georgia indicate that the assessment process and the resultant profile have helped them identify their strengths and weaknesses. Over 90% indicate that they have used the results to improve their teaching.

It is critical that help be provided in areas of **assessed** need, not merely felt or perceived need. This assistance can and should take many different forms and must be the least structured of all procedures involved in the development of teachers. Because of the individual behavioral changes that take place throughout the assessment process, not all teachers will need, or even profit from, many formal staff development programs. However, for some, such programs are vital; for most, they enhance and reinforce teaching competencies.

Staff development opportunities for beginning teachers in Georgia are based on analyses of needs revealed in the individual profiles. Programs include a resource guide keyed to the competencies, released time for beginning teachers and peers, and courses for graduate credit or staff development units related to on-the-job needs. The state provided these resources to local school systems as they developed plans for beginning-teacher staff development. Institutions of higher education can provide staff development through field-based programs, as well.

Prior to the comprehensive process, only one out of eight Georgia teachers could demonstrate mastery of the essential competencies; now, more than four out of five master the competencies and over 90% of those assessed say that they are proud of the essential on-the-job competencies they are able to demonstrate.

In Georgia, the responsibility for providing staff development opportunities to teachers rests with the local school system. However, in many areas of the state, staff development opportunities are available from a Cooperative Education Service Agency (CESA). Regional assessment centers are not involved in staff development.

4. EVALUATION FOR PROMOTIONS AND SALARY INCREASES

Georgia's experience with its performance-based teacher certification system provides some valuable insights for policy makers, planners and implementers of career-ladder or similar programs. In a few years, the early experience with new teacher incentive programs in "pioneer" states like Florida and Tennessee will enrich the bases for decision making, planning and implementing.

For the time being, planners will have to rely on a limited experience base, continuing research, a common-sense approach to what is likely to work under what kinds of conditions and their own careful experimentation. The territory may be less treacherous if planners have a firm grasp on what it is they know, and can identify with some certainty what it is they don't know. Unknown factors may include these:

- **Do *outstanding* teachers have specific characteristics and competencies that average teachers do not display? What are they? How can they be measured? Is there a way to pass them on to average teachers? Can they be taught in preservice training institutions?**
- **Will teachers accept the limited — and competitive — nature of plans like career ladders? These programs necessarily call for extra funding, not only for monetary rewards, but for carrying out evaluations and staff development efforts. Participation likely will be limited. Will teachers unable to qualify for such systems be content with what they have? Can these systems be provably fair enough to turn aside teacher disaffection and disgruntlement?**
- **Will high-performance teachers who choose not to participate in such systems be comfortable with their status? Should planners develop a different kind of recognition for these valuable educators?**
- **What kinds of continuing challenges — if any — should be offered to teachers at the top of the heap? Can schools afford to lose them through the hole at the top (promotion to administrator, for example) or through the hole at the bottom (boredom, burnout)?**
- **Will public identification of outstanding teachers lead to public dissatisfaction with average — and *adequate* teachers? ("My son has a superior teacher; my daughter's teacher is only average.") How can this possibility be averted?**
- **Is evaluation feedback really helping teachers to improve their classroom performance? Does the information they receive about performance truly challenge them to improve? Are performance expectations realistic and attainable? Do support systems to help teachers improve truly match their needs?**

REFERENCES

Braskamp, Larry A. et al. **Evaluating Teaching Effectiveness: A Practical Guide.** Beverly Hills: Sage Publications, Inc., 1984.

Cresap et al. **Teacher Incentives: A Tool for Effective Management.** Reston, Va.: National Association of Secondary School Principals, 1984.

Darling-Hammond, Linda and Arthur E. Wise. "Testing Teachers for Certification: An Overview of the Issues," a paper prepared for the Minnesota Higher Education Coordinating Board, August 1984.

Fullan, Michael. **The Meaning of Educational Change.** New York: Teachers College Press, Columbia University, 1982.

Goertz et al. **The Impact of State Policy on Entrance into the Teaching Profession,** excerpts on California and Georgia. Princeton, N.J.: Educational Testing Service, October 1984.

Hatry, Harry P. and John M. Greiner. **Issues in Teacher Incentive Plans.** Washington, D.C.: The Urban Institute, 1984.

House, Ernest. **Evaluation With Validity.** Beverly Hills: Sage Publications, 1980.

Joyce, Bruce R. and Beverly Showers. **Power in Staff Development Through Research on Training.** Alexandria, Va.: Association for Supervision and Curriculum Development, 1983.

Lewis, Anne C. **Evaluating Educational Personnel.** Arlington, Va.: American Association of School Administrators, 1982.

McGreal, Thomas L. **Successful Teacher Evaluation.** Alexandria, Va.: Association for Supervision and Curriculum Development, 1983.

Millman, Jason, ed. **Handbook of Teacher Evaluation,** published in cooperation with the National Council on Measurement in Education. Beverly Hills: Sage Publications, Inc., 1981.

Wise, Arthur E. et al. **Teacher Evaluation: A Study of Effective Practices,** R-3139-NIE. Santa Monica: The Rand Corporation, June 1984.

HOW TO ORDER BOOKS IN THIS TEACHER QUALITY SERIES

Release dates for this nine-booklet series will begin November 1 through December 31, 1984. For more information about the issues discussed in these booklets, write or call Robert Palaich at the ECS Denver address, 303-830-3642.

Booklets are priced at $6 each; a full set will be offered at $36. For ordering information or to find out which booklets are available as they are produced over the next few months, write or call the ECS Distribution Center, 303-830-3692.

Booklets are described below. Please use both number and title when ordering.

1. **A Policy Guide to Teacher Reward Systems** by Ellen Flannelly and Robert Palaich, Education Commission of the States, TQ84-1

   The authors present brief arguments for and against major positions on selecting goals for performance pay systems, setting performance standards, designing evaluation programs, different kinds of pay systems and other ways to improve teaching. They also offer a bibliography to support their arguments.

2. **Evaluating Teachers: With Lessons From Georgia's Performance-Based Certification Program,** by Doris Ross, Education Commission of the States, and Lester M. Solomon, Georgia Department of Education, TQ84-2

   Solomon, writing out of his experience in designing and carrying out a pioneer teacher evaluation plan in Georgia, overviews evaluation procedures accompanying performance-based pay and staff development, and compares testing and on-the-job assessment. He recommends appropriate timing, outlines how to use tests to establish minimum competencies, describes methods of training evaluators and warns against expecting more than evaluation techniques can deliver.

3. **Improving Teacher Quality Through Incentives** by Robert Palaich and Ellen Flannelly, Education Commission of the States, TQ84-3

   Palaich and Flannelly suggest ways for policy makers to clarify their goals for reward-for-performance plans so they may select the most appropriate plans. They set limits on expectations for monetary incentive plans by discussing research that shows that teachers are strongly influenced by intrinsic motivation, school organization and interaction with colleagues, as well as by money. They point out that plans must include clear performance standards and evaluation systems, and that both evaluators and teachers must be trained to use them. Finally, they offer models of merit pay, career ladders and personnel distribution incentives.

4. **Political Myths About Reforming Teaching** by Susan J. Rosenholtz, Vanderbilt University, TQ84-4

   Ten common beliefs about how performance-based pay and promotions will help improve teaching are compared to research findings in this book, and the author concludes that they don't hold up. Although low pay discourages the academically able from entering or remaining in teaching, the author presents research that shows teachers to be more frustrated by their lack of success with students. Rosenholtz identifies the conditions that support effective teaching, states that almost all teachers can improve, cautions against using student test scores as measures of teaching effectiveness and warns that competition for rewards among teachers may mitigate against essential collaboration among teachers and administrators.

5. **State Strategies To Improve Teaching** by Robert Palaich, Education Commission of the States, TQ84-5

   Local efforts to improve teacher quality can be initiated and/or bolstered by state

actions, and Palaich offers a logical cumulative strategy for these actions. He covers screening for admission to schools of education, improving curriculum, graduation requirements, certification and tenure. He also shows how states can help develop and fund better evaluation systems, in-service training and performance reward systems, explaining that certain areas of choice should be left to local districts.

6. **The Legal Context for Teacher Improvement** by the Education Commission of the States' Law and Education Center, TQ84-6

   In an effort to pre-inform policy makers and administrators contemplating teacher improvement plans, ECS Law Center staff explain the legal aspects that may affect these plans, and discuss how to tailor plans to comply with constitutional and statutory requirements. Due process, civil rights, free speech, academic freedom, tenure, collective bargaining and governance issues are covered. Case cites and a selected bibliography support the authors' arguments.

7. **A Guideline for Evaluating Teacher Incentive Systems** by Steven M. Jung, American Institutes for Research, TQ84-7

   Jung develops a conceptual framework for evaluating teacher incentive systems. A performance-based system, he says, bases rewards on behavior rather than on added responsibilities. Stated goals must mesh with goals in practice if evaluations are to be valid. Jung also examines assumptions about teaching excellence and the process components of incentive systems.

8. **School Organization and the Rewards of Teaching** by Thomas Bird, Boulder, Colorado, TQ84-8

   Bird focuses on how to organize schools and school settings to encourage better teaching. He describes organizational schemes that encourage staff to share understandings and techniques, help each other to improve and use research findings to test new methods. He suggests that teachers and administrators be trained as role models, and recommends that experimental research applications be supported at the state level.

9. **The Costs of Performance Pay Systems** by Kent McGuire, Education Commission of the States, and John A. Thompson, University of Hawaii, TQ84-9

   Using two different evaluation systems, the authors simulate the costs of merit pay, career ladders and extended contracts to show how costs — none of them prohibitive — vary with plan design. The authors precede the simulations with a thorough discussion of each cost factor involved.

Designer: Patricia J. Kelly
Typographer: Marci Reser
Printer: Golden Bell Press, Denver